Towards an Understanding of Human Persuasion and Biases in Argumentation (CAF 2016)

Pierre Bisquert, Madalina Croitoru, Florence Dupin de Saint-Cyr, Abdelraouf Hecham

▶ **To cite this version:**

Pierre Bisquert, Madalina Croitoru, Florence Dupin de Saint-Cyr, Abdelraouf Hecham. Towards an Understanding of Human Persuasion and Biases in Argumentation (CAF 2016). Cardiff Argumentation Forum (CAF 2016), Federico Cerutti (School of Computer Science & Informatis at Cardiff University); Francesca Toni (Department of Computing at Imperial College London), Jul 2016, Cardiff, United Kingdom. hal-03155070

HAL Id: hal-03155070
https://hal.science/hal-03155070

Submitted on 1 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards an Understanding of Human Persuasion and Biases in Argumentation

Pierre Bisquert, Madalina Croitoru, Florence Dupin de Saint-Cyr, Abdelraouf Hecham

GraphIK, IATE, INRA, Montpellier, France

GraphIK, LIRMM, University of Montpellier, France

IRIT, University Paul Sabatier, France

Abstract

We present in this paper some recent work aiming at allowing the formal analysis of the persuasive impact that an argument may produce on a human agent based . We present a computational model based on the Dual Process Theory and applied to argument evaluation. These works form the preliminary step that will allow a better understanding of two crucial aspects of collective decision-making: persuasive processes and argumentation strategies.

Introduction

Gaining more and more attention, persuasion is a crucial aspect of human interaction. With the recent rise of computer science technology, the study of persuasion began to transcend its original fields (including psychology, rhetoric and political sciences) and to take lasting root in the artificial intelligence domain. We are interested in the link between persuasion and cognitive biases, and why well-founded arguments are rejected or fallacious arguments are accepted

The highly influential cognitive psychology work in dual systems ([11, 3, 6, 1, 5, 9]) associate such biases with two reasoning systems: one system that is slow but logically precise and another system that is fast but logically sloppy. The distinction does not make clear the interaction between biases due to logically flawed reasoning and biases due to suboptimal reasoning choices done because of cognitive limitations. This distinction is interesting and useful to consider when addressing the evaluation of biased argument.

We propose to investigate the problem of argument evaluation by agents that are logically biased (i.e. may either reason exclusively logically or by combining logical reasoning with associations).

In our proposal we follow and hypothesize that, when it is not possible for an agent to make a logical inference (too expensive cognitive effort or not enough knowledge), she might replace certain parts of the logical reasoning with mere associations (such systems are also known as dual process systems in cognitive science literature). Using associations may lower the reasoning effort needed for argument evaluation and subsequently affect the argument acceptance.

Proposal

Standard Artificial Intelligence work in computational cognitive systems has never considered looking in depth at the distinction between machine agents and human agents, even more analyse human reasoning. This is largely due to the fact that Artificial Intelligence has been long influenced by the view that logic is the theory of ideally good thinking desired of intelligent agents (human or artificial).

In practice such good thinking is often far from reality. The harsh reality is particularly illustrated in the human agent interaction paradigm applications where human irrationality (biases) is a major problem (see [10] for an overview).

In artificial agents two kinds of biases were highlighted by existing literature ([4], [7], [8]). On one hand, the agent's beliefs and preferences may be incomplete and agent may not know all preferences or beliefs needed for complete reasoning. This kind of representational issued biases could be linked to the so-called Type 1 irrationality or substantive irrationality ([11, 3, 6, 1, 5, 9]) that concerns the compliance of the results of reasoning with the agents knowledge base. In the proposed line of work, analyzing

this irrationality amounts to know if the agent has used logical inference to perform the reasoning or, if the reasoning has been also partially (or totally) performed on associations.

Procedural flaws of reasoning concern the case when, due to the fact that computational resources (time or space available for representing and reasoning) are limited, the agent needs to make good choices in the process of deciding how to apply its efforts in reasoning. We argue that achieving procedural rationality means making rational choices about what inferences to perform, how to apply them, basically thinking about how to think.

The difference between the two kinds of reasoning errors is important to distinguish. The first case verifies if the results of reasoning satisfy the agent knowledge base, while the second case verifies if the agents makes good choices in the process of deciding how to apply its best efforts for reasoning. So far, in argumentation literature, the two kinds of biases are not clearly distinguished. Existing work either addresses substantive biases in a Kahneman inspired system for propositional logic ([2]) or some bias inspired reasoning procedures (procedural rationality) mainly related to persuasion approaches ([12]).

We will study the properties of the possible reasoning paths that an agent can follow. This is why we might assume that it is possible to compute a multiplicity of reasoning paths and compare them, while in real life it is probable that a human agent will not do so. It is important to note that we do not define how a human being reasons but we try to obtain the preferred reasoning paths that a human could obtain.

References

[1] Christopher G. Beevers. Cognitive vulnerability to depression: A dual process model. Clinical Psychology Review, 25(7):975 – 1002, 2005.

[2] Pierre Bisquert, Madalina Croitoru, and Florence Dupin De Saint Cyr. Towards a dual process cognitive model for argument evaluation. In Christoph Beierle and Alexander Dekhtyar, editors, International Conference on Scalable Uncertainty Management (SUM), Quebec, 16/09/2015-18/09/2015, LNAI, pages 298–313, 2015. Springer.

[3] Pat Croskerry, Geeta Singhal, and Salvia Mamede. Cognitive debiasing 1: origins of bias and theory of debiasing. BMJ Quality & Safety, 22(Suppl 2):58–64, 2013.

[4] Jon Doyle. Rationality and its roles in reasoning. Computational Intelligence, 8(2):376–409, 1992.

[5] Seymour Epstein. Integration of the cognitive and the psychodynamic unconscious. American Psychologist, 49(8):709, 1994.

[6] Jonathan St. B. T. Evans and Jodie Curtis-Holmes. Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. Thinking & Reasoning, 11(4):382–389, 2005.

[7] IJ Good. The probabilistic explication of information, evidence, surprise, causality, explanation, and utility. Foundations of statistical inference, pages 108–141, 1971.

[8] Herbert A Simon. From substantive to procedural rationality. In 25 Years of Economic Theory, pages 65–86. Springer, 1976.

[9] Steven A. Sloman. The empirical case for two systems of reasoning. Psychological Bulletin, 119(1):3, 1996.

[10] Joseph Douglas Terwilliger and Jurg Ott. Handbook of human genetic linkage. JHU Press, 1994.

[11] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. Science, 185(4157):1124–1131, 1974.

[12] Frans H Van Eemeren and Rob Grootendorst. Argumentation, communication, and fallacies: a pragma-dialectical perspective. Lawrence Erlbaum Associates, Inc, 1992.